Available online at www.sciencedirect.com

SciVerse ScienceDirect

IERI Procedia 2 (2012) 543 – 547

Procedia IERI

www.elsevier.com/locate/procedia

2012 International Conference on Future Computer Supported Education

# Design of Publishing Information Service System Based on Web3.0

Huiying Qi[a], Yazi Li[b]*

[a]*Medical Science Department, Peking University Health Science Center, Beijing,100191, China*
[b]*Institute of Medical Information, Chinese Academy Of Medical Sciences, Beijing,100020, China*

**Abstract**

Due to the shortcomings of web2.0 technology lacking efforts to support the discovery process of scientific research and reuse of existing products, a complete set of universal semantic web-based publishing information service framework is proposed based on the currently successful case of digital publishing architecture and technology. The publishing information service system established under the framework can automatically extract key concepts of the resource content.



Selection and peer review under responsibility of Information Engineering Research Institute

*Keywords:* publishing information service system, web3.0,semantic web, ontology;

## 1. Introduction

A certain extent, the application of web2.0 technology enhances the content quality and diffusion rate of publishing information service. But it lacks efforts to support the discovery process of scientific research and reuse of existing products. The emergence and development of web3.0 changes the publishing process involved in various aspects: facilitating the readers, expanding the publisher production situation, changing the method of content organization, increasing content distribution efforts, making the computer understand contents of the article, mining the tacit knowledge and promoting the new findings[1] [2].

* Corresponding author. Tel.: 8610-10-52328871.
*E-mail address:* li.yazi@imicams.ac.cn.

 doi:10.1016/j.ieri.2012.06.131

Ideal semantic web-based information service system involves to promoting content --- including a wide range of data and knowledge generated by the publication process, full text of journal indexing, keywords and other records of the content. It's able to connect to the book content, laboratory sites, factual data and high-level comments.

Vast amounts of information are interlinked, and users can access these linked data through the one-stop portal. To accomplish these functions, a series of standards should be developed, using the semantic web to describe resources and services, achieving the release of contents of publishing information service system. It can effectively achieve interoperability, integrate resources and service and support data mining [3] [4].

## 2. Current Research

Semantic web technologies make the published content releasing and receiving more and more convenient to use and showing intelligence. The implementation of these technologies has become a part of the publishing industry. Now publishers are positively rolling out the interpretation of linked data, data mining, intelligent retrieval and data analysis services for the content [5] [6]. For example, Chinese Publishing House developed *Comprehensive Mirror for Aid in Government Analysis System*[7]. The system is based on ontology and semantic analysis technology, through the text processing and knowledge recombining to build an open knowledge service system. It realizes the knowledge-based content retrieval and knowledge discovery used by semantic analysis technology. It is a new tool of reference, retrieval and using antique book.

In the field of press and publication, it focuses on news events including time, place, characters, causes, event results and other core information. Using of semantic web technologies, any of news can be made into a number of special topics. Subject content of the special topic is not limited the site, and hot news of other sites can be linked automatically. For example, *The Wall Street Journal* ( http://www.wsj.com) introduces U.S. News, European news, Asian news, technology news, business news, stock fax, expert commentary and so on. Semantic web technologies associate with news events, people, events and the trend of events. The value of information is greatly increased [8].

In addition, the academic site StemBook (www.stembook.org) based on domain ontology provides accurate explanations and instructions to a variety of technical terms in the articles. And it can be extended unlimited class and release open links from articles to related resources of a variety of disciplines or areas [9].

## 3. System Architecture

Semantic-based publishing information service system supports the expansion of the publishing process and the core data index. It keeps the contribution that scientists spread of the scientific and cultural knowledge. The architecture is divided into four levels: resource layer, annotation layer, index layer and user layer. It is showed in figure 1.

### *3.1. Resource Layer*

Resource layer consists of bibliographic information of articles (such as abstracts, titles, authors and publishers), other units of structured information (such as tables and pictures in the articles), experimental data sets, the articles in the involved subjects as well as to achieve the results obtained related to the data set. Resource layer provides abstracts and full texts published by the key elements of digital content. Public API can be used to retrieve the text of these elements metadata resources, and track the original elements and elements id.

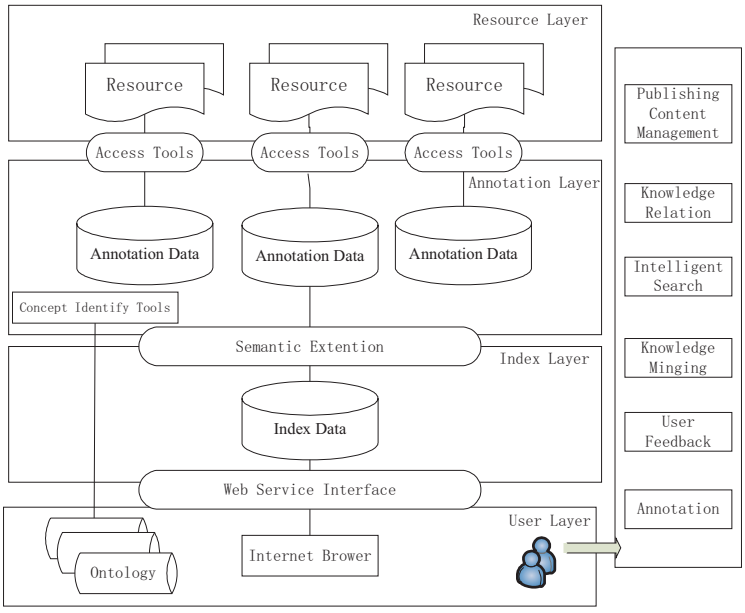


Fig. 1. Semantic -based Publishing Information Service System Architecture Diagram

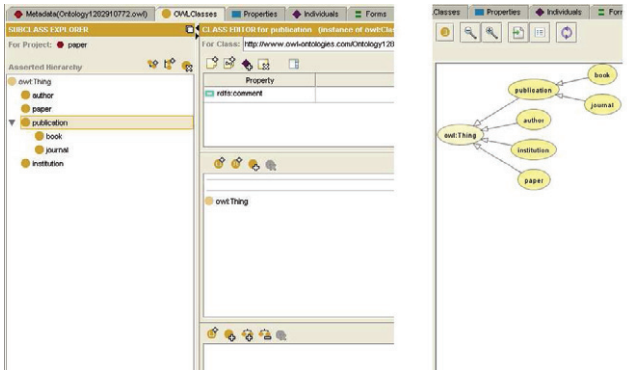

Fig. 2. RDF concept space

Resource layer is shown as follows:

- RDF

The resources are organized in RDF language, such as journals, papers, authors, institutions, citations and so on. This is an example that papers build a simple RDF concept space. Figure2 shows four types of knowledge: journal (which is a subclass of publication), articles, authors and institutions.

- XML and XHTML

The system provides the XML format. It facilitates third-party systems through a common interface calling.

It is displayed to the end user in XHTML format. The information includes the journal information and the paper information (such as title, author, full-text, pictures, tables and citations). It connects between the outside resources and the element extracted from the annotation layer.

### *3.2. Annotation Layer*

Annotation layer can be quickly retrieved related concepts of the body. Text content can be annotated by the word or concept of dictionary. After annotated, it is to track title and summary using the annotation content. The result is often stored in the annotation table.

### *3.3. Index Layer*

In the index level, the global index associates all index tables and indexes by the concept of ontology indexing. For example, the concept of T marked elements E1, E2.

### *3.4. User Layer*

In the user layer, the result can be marked directly by the resource elements when searching for specific ontology concept. The users accept the search results according to reference and link. The user layer includes the following key features:

- Published Content Management

  Published content management improves the author's collaboration efficiency. In the semantic context, custom filtering policies achieve modelling based on user requirements.

  Editorial is able to find the right information associated Reviewer, a standard data description format of the organization publishing the content, and enhance their interoperability, while allowing readers to view the content, thereby enhancing the magazine's influence.

- Knowledge Relation

  Digital publishing covers a lot of books, periodicals, audio, video and other unstructured data. Knowledge points scattered in various carriers can be automatically extract through the establishment of the ontology. Ontology associates with the different point of publishing content and extracts the most basic unit of knowledge content based on data resource layer. This has played a supporting role to manual labor.

- Knowledge Cutting

  The digital publishing content management system can focus word, sentence, paragraph and chapter based on the important points of user's reading papers and books. It highlights the important contents of papers and users interested in. The content-based knowledge point cutting can associate with related knowledge points, and expand the knowledge content of published content itself. It is convenient for the readers.

- Intelligent Search

  Using intelligent retrieval, the users input natural language. Then the system automatically converts into the entry word of the ontology and positions the concept of entrance word. This achieves concept-based queries, which can display the full knowledge points. The search results return with automatic classification and clustering based on user input query. At the same time, it supports user-defined classification.

- Knowledge Mining

  By building a model based on user needs, it can mine the useful information and services according to the needs of users, such as recommended the latest research of subject areas for research users, and organized the related article's background knowledge or well-known author of the field for the authors. For publishers, the adoption of content clustering can know the publication's column and understand the structure of the knowledge evolution. It is also improved the editorial and review process efficiency and effectiveness.

## 4. Conclusion

The emergence and development of Web3.0 technology change the publishing process and involved roles in the publishing process. It facilitates the readers, expands the production situation, changes the method of content organization and increases content distribution efforts. It also makes the computer understand the contents of the article, taps the tacit knowledge. The Web3.0 technology provides a strong method for promoting the new findings.

## References

[1] Louise C,Nigel S. Extracting focused knowledge from the semantic web. Int. Human-Computer Studies, 2007;54:155-184.
[2] Jason J. Exploiting Semantic Annotation to Supporting User Browsing on the Web. Knowledge-Based Systems , 2007;5:373-381.
[3] Fensel D. The Semantic web and its languages. IEEE Computer Societu, 2000;11-12:67-73.
[4] Liu BS,Gao J. General ontology learning framework. Journal of Southeast University ,2006; 22:381-384.
[5] Clark T, Kinoshita J. Alzforum and SWAN: the present and future of scientific web communities. Briefings in Bioinformatics,2007;8:163-171.
[6] Hignette G, Buche P,Couvert O,et al. Semantic Annotation of Web Data Applied to Risk in Food. International Journal of Food Microbiology,2008; 11:174-180.
[7] Comprehensive Mirror for Aid in Government Analysis System. http://www.cnpubg.com.cn/zt/newsdetail.cfm?iCntno=63051,2012-05-08.
[8] The Wall Street Journal. http://www.wsj.com, 2012-05-08.
[9] About StemBook. http://www.stembook.org/about, 2012-05-08.